User QoE Assessment on Mobile Devices for Natural and Non-natural Multimedia Clips

Arghir-Nicolae Moldovan, and Cristina Hava Muntean
School of Computing, National College of Ireland, Mayor Street, IFSC, Dublin 1, Ireland
E-mail: arghir.moldovan@ncirl.ie; cristina.muntean@ncirl.ie

***Abstract*—Quality of Experience (QoE) has become an increasing topic of research with the proliferation of multimedia services on mobile devices. Previous research studies have focused on proposing objective video quality assessment (VQA) metrics, and evaluating them on generic video content databases that consist mainly of natural clips such as news, sports and movies. This paper investigates the accuracy of VQA metrics to estimate user's QoE for natural and non-natural multimedia clips on mobile devices. The results from a subjective study with 60 participants have shown that well known full-reference VQA metrics such as PSNR, SSIM and VIFp exhibit up to 97% QoE estimation accuracy for non-natural clips, despite not being traditionally recommended for such clips.**

***Index Terms*—Quality of Experience (QoE), Video Quality Assessment (VQA), objective metrics, continuous quality estimation.**

I. Introduction

With the mobile technologies global adoption on a fast rise, Internet and computing are arguably in the middle of a mobile revolution age. Mobile devices such as as smartphones and tablets are getting more powerful and complex, but also low cost and increasingly popular among users. These devices have become mobile work, learning and entertainment centres being used for a multitude of activities including resource intensive multimedia applications.

Online services that involve multimedia streaming to mobile devices have been growing at a fast pace in recent years, thanks to the advancements in video and networking technologies and due to the online shift of TV, movie and video content. Cisco predicted that mobile video will increase 8-fold between 2015 and 2020, accounting for 75% of total mobile data traffic by the end of 2020 [1]. The challenge for mobile service providers is both to effectively manage the high volume of video traffic that smartphone users generate, and to fulfil end-users' expectations in terms of having access to high-quality video content and rich multimedia applications.

As mobile users become more quality-aware, there is a growing need for automatic and reliable metrics to estimate users' QoE with multimedia services [2]. While a multitude of objective VQA metrics have been proposed [3], previous research studies have primarily focused on evaluating their performance on generic video content databases that consist mainly of natural clips such as news, sports and movies [4]. As many metrics are based on visual statistics and features of natural scenes, non-natural clips were often excluded as they were considered to complicate the metrics' evaluation [5]. Natural clips correspond to video recordings of real-world scenes, while non-natural clips are usually computer-generated.

This paper investigates the accuracy of well-known VQA metrics such as PSNR, SSIM and VIFp to estimate the user QoE measures resulted from perceptual testing on mobile devices. The main contribution of the paper consists in the analysis of non-natural clips along with natural clips. A subjective study was conducted with 60 participants that had to continuously rate their perceived QoE for 6 educational multimedia clips with changing quality level. The results showed that VQA metrics can estimate the user-perceived quality of non-natural educational clips with up to 97% accuracy.

The rest of the paper is structured as follows. Section II presents related work in the area of video quality assessment. Sections III and IV present the setup and results analysis for the subjective study, while section V concludes the paper.

II. Related Work

The area of video quality assessment has seen much research work and interest from both the academia and industry over the years. The video quality can be evaluated using subjective methods and objective metrics.

Subjective methods are considered the most accurate and reliable way for assessing the video quality. A number of methods were standardised by ITU in the recommendations ITU-R Rec. BT.500 [6] for television and ITU-T Rec. P.910 [7] for multimedia applications. These standards provide useful guidelines and instructions regarding the selection of the subjects and of the test material, the setup of the test environment, the rating scales to be used for assessment, as well as the methods for analysing the data. Subjective methods can differ in many aspects such as the test sequences presentation (i.e., double stimulus vs. single stimulus), the rating moment (e.g., after viewing short sequences vs. continuously while viewing long sequences), or the rating scale (e.g., discrete 1 – bad to 5 – excellent scale vs. continuous 1–100 scale). While MOS scales are often criticised to not accurately indicate users' QoE in terms of acceptability, Spachos *et al.* [8] have shown that the technical quality and overall experience measured on a 5-point scale impact on the user acceptability expressed as a binary

This research is supported by the NEWTON project (http://www.newtonproject.eu/) funded under the European Union's Horizon 2020 Research and Innovation programme, Grant Agreement no. 688503.

978-1-5090-1990-8/16/$31.00 ©2016 IEEE

measure. The main limitation of subjective methods is their reduced applicability for real-world multimedia applications as they are cost and time expensive due to the need for participants to provide their opinion

Objective VQA metrics aim to provide automatic quality estimation and are more suitable for real-world applications. The objective VQA metrics can be classified in: full-reference (FR), reduced-reference (RR) and no-reference (NR) metrics. FR metrics provide the highest quality estimation performance, but require precise spatial and temporal synchronisations between the original and impaired videos. Peak Signal-to-Noise Ratio (PSNR) provides a baseline for video quality assessment metrics and continues to be widely used thanks to its simplicity. More complex metrics such as Structural Similarity Index (SSIM) [9] and Visual Information Fidelity (VIFp) [10], are based on natural visual characteristics or models of Human Visual System (HVS).

RR metrics (e.g., [11], [12]) aim to provide a compromise between accuracy and flexibility, by making use of some information extracted from the reference video, such as the amount of motion or spatial detail, which are more feasible to be transmitted over the communication channel.

NR metrics (e.g., [13], [14]) have a high flexibility since they do not require the presence of the original video, unaffected by the factors under test. The majority of these metrics estimate the video quality of the impaired clip based on factors such as blockiness, blurring, jerkiness, ringing, etc.

A major challenge is to decide between the multitude of objective metrics, especially for applications such as multimedia mobile learning that often involve non-natural clips, which were traditionally excluded from subjective evaluation studies.

III. Subjective Study Setup

A subjective study was conducted in order to evaluate the performance of the PSNR, SSIM and VIFp full-reference objective VQA metrics to estimate the continuous QoE for educational multimedia clips on mobile devices. The three metrics were selected as they correspond to different categories (i.e., traditional point-based, natural visual statistics-based, and perceptual HVS modelling-based [3]).

A. Multimedia Test Sequences

Six high-quality clips corresponding to different categories of educational multimedia clips were used for the subjective study. ArtOfBook (documentary), NitrogenIceCream (demo), and ProjectPlanning (presentation) are mainly natural clips, while AtomSize (animation), CoralsIntro (slideshow), and PhotoEditing (screencast) are mainly non-natural clips. More details about the clips can be found in [15].

A 4 min long continuous test sequence was extracted from each educational clip. Fig. 1 presents the spatio-temporal complexity of the test sequences quantified through the Spatial Index (SI) and Motion Vectors (MV) metrics computed as in [4]. Each sequence was compressed using the H.264 video codec at 1 reference and 13 other quality levels with different

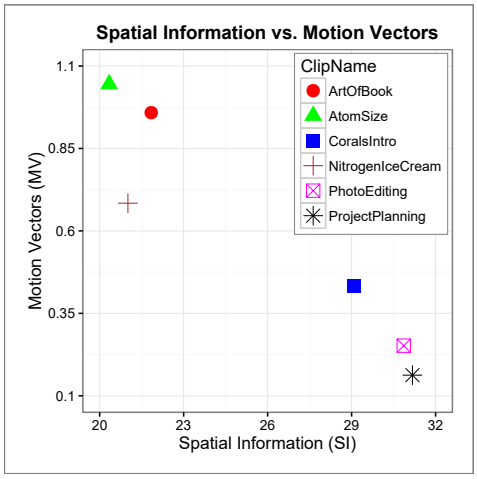

Fig. 1. Spatio-temporal complexity of the multimedia test sequences.

TABLE I
Encoding settings for the testing scenarios.

	⟨**Resolution** [pixels], **Framerate** [fps], **Bitrate** [kbps]⟩		
15s Int.	**Scenario 1 (S1)** (***Resolution Decrease***)	**Scenario 2 (S2)** (***Framerate Decrease***)	**Scenario 3 (S3)** (***Bitrate Decrease***)
I0	⟨**1280**×**720**, 30, 1800⟩	⟨1280×720, **30**, 1800⟩	⟨1280×720, 30, **1800**⟩
I1	⟨**1280**×**720**, 30, 1500⟩	⟨1280×720, **30**, QP25⟩	⟨1280×720, 30, **1024**⟩
I2	⟨**848**×**480**, 30, 600⟩	⟨1280×720, **15**, QP25⟩	⟨1280×720, 30, **384**⟩
I2	⟨**640**×**360**, 30, 350⟩	⟨1280×720, **7.5**, QP25⟩	⟨1280×720, 30, **128**⟩
I4	⟨**426**×**240**, 30, 150⟩	⟨1280×720, **3.75**, QP25⟩	⟨1280×720, 30, **64**⟩
I5	⟨**320**×**180**, 30, 90⟩		

values for the video bitrate, framerate and resolution parameters (see section III-B for the specific values used). Apart of the three parameters all other video and audio encoding settings were maintained constant.

Following that, consecutive 15 sec long segments were extracted from the 14 different quality versions and joined together to obtain the test sequences with changing quality. The AviSynth [16] nonlinear video editor was used for performing on-the-fly video editing tasks on the original clips (i.e., trimming/joining sequences, changing the resolution or framerate), without the need for recompression. More details on how the test sequences were created can be found in [17].

B. Test Scenarios

Three testing scenarios were considered for the study, which consisted of gradually decreasing the three video encoding parameters: resolution (S1), framerate (S2) and bitrate (S3). Table I presents the encoding settings for the different 15 sec intervals of each scenario. The encoding values for the reference quality (I0), and the bitrate values for the different resolutions were selected based on the educational multimedia profiling recommendations from [18]. For S2 the compression level was maintained constant at QP = 25 quantization factor.

The three scenarios were conducted in succession over the 4 min duration of each test sequence, but in a different order (e.g., S3-S2-S1, S1-S3-S2, etc.). The reference quality level

Fig. 2. On-screen slider used for the continuous video quality rating.

was used for the first 15 sec interval of each test scenarios in order to provide a baseline and help the participants re-adjust their opinion. Using long test sequences with changing quality level enables a real-world like multimedia viewing experience.

C. Subjective Testing Procedure

The subjective study was conducted with 60 non-expert volunteer participants (37 males, 23 females), aged between 20 to 53 years old ($AVG = 28.67, SD = 6.87$). The participants were asked to view each of the 6 multimedia test sequences with changing quality level and continuously rate their perceived video quality. The standardised Single Stimulus Continuous Quality Evaluation (SSCQE) procedure was followed, with rating done on a calibrated 0 – 100 continuous scale with annotated QoE levels (i.e., 'bad' to 'excellent') [6]. The rating was done using an on-screen slider displayed over the video player, as presented in Fig. 2. The viewing and rating of the test sequences was done on a Google Nexus 7 tablet mobile device, with a 7 inch screen and 1280×800 resolution.

To counteract any effects such as fatigue on the subjective results multiple randomisations were performed (i.e., testing scenarios randomisation across educational clips, and educational clips randomisation across participants).

IV. Results

A. Continuous MOS and VQA Metrics Analysis

Fig. 3 presents the continuous MOS (Mean opinion Score) across the 60 participants averaged for each second interval. The figure also illustrates the objective values for the VIFp metric computed between the reference quality version and the changing quality version of each test sequence, averaged on a per second basis (i.e., across every 30 frames). The equivalent QoE bands on the 5-point discrete scale (i.e., 1 – 'bad' to 5 – 'excellent'), and corresponding VIFp thresholds based on the mapping solution proposed in [19], are also illustrated.

The continuous MOS results show that the participants noticed the decrease in quality for all 3 test scenarios, but usually there is approximately 5 sec delay until they adjust the slider. The results also show that the QoE decreases higher in case of the resolution (S1) and bitrate (S3) decrease scenarios. For the framerate (S2) decrease scenario the participants rated as 'excellent' or 'good' the sequences even at 3.75 fps. The QoE decrease is related to the clip characteristics. For example, the PhotoEditing clip presents high spatial and low temporal detail (see Fig. 1), thus the QoE decreases to 'poor' in case of S1 and S3 but remains within 'excellent' for S2.

The continuous VIFp results show that while the metric captures the decreasing QoE trend, it presents very high variations. This is expected as the VQA metrics are computed on a per-frame basis. However, as the human participants do not notice the changes in quality at such high granularity it is safe to average the values across multiple seconds.

Fig. 4 presents the MOS and VIFp values averaged for each quality level (i.e., each 15 sec interval). The results show that VIFp can capture with high accuracy the decreasing QoE trend, as well as the QoE level, especially for S1 and S3.

B. Quality Estimation Accuracy Results

Figs. 5 and 6 present the quality estimation accuracy results quantified using the Pearson Linear Correlation Coefficient (PLCC) [3]. The PLCC measure was computed after non-linear regression between the MOS and VQA metric values using a cubic polynomial function as in [19]. PLCC was computed separately for each group of clips (i.e., natural vs. non-natural), and for each of the S1, S2, and S3 test scenario in three cases: continuous MOS (and VQA metric values) averaged over 1, 7.5 and 15 second intervals.

The results show a number of interesting findings:

- The quality estimation accuracy increases if the MOS and VQA metric values are averaged over a longer interval.
- All three VQA metrics perform better for natural than non-natural clips. However, PSNR and VIFp perform almost equally well independent of content type for S1 and S3, while SSIM performs significantly worse for non-natural clips on S3.
- The VQA metrics tend to present lower performance for S2 framerate decrease. However, the poorer performance is manifested to higher a degree in case of non-natural clips, with maximum 67% accuracy provided by PSNR on 15 sec average, while for natural clips the metrics provide over 79% accuracy in all test cases.

V. Conclusions

This paper investigated the performance of well-known VQA metrics such as PSNR, SSIM and VIFp to estimate user's QoE for different categories of educational multimedia clips, including natural (i.e., documentaries, presentations and demos), and non-natural (i.e., animations, slideshows and screencasts). The investigation on mobile devices was conducted in a continuous way with long sequences to provide a more realistic experience.

The results analysis has revealed that user QoE is affected to a higher degree by a decrease in resolution or bitrate, as compared to decreasing the framerate. The PSNR, SSIM and VIFp metrics are more suitable to estimate the user QoE for natural clips as the accuracy was higher than 79% for all test conditions. Moreover these metrics also perform very well for non-natural clips with up to 96% accuracy for resolution decrease, up to 97% for bitrate decrease, but only up to 67% for framerate decrease.

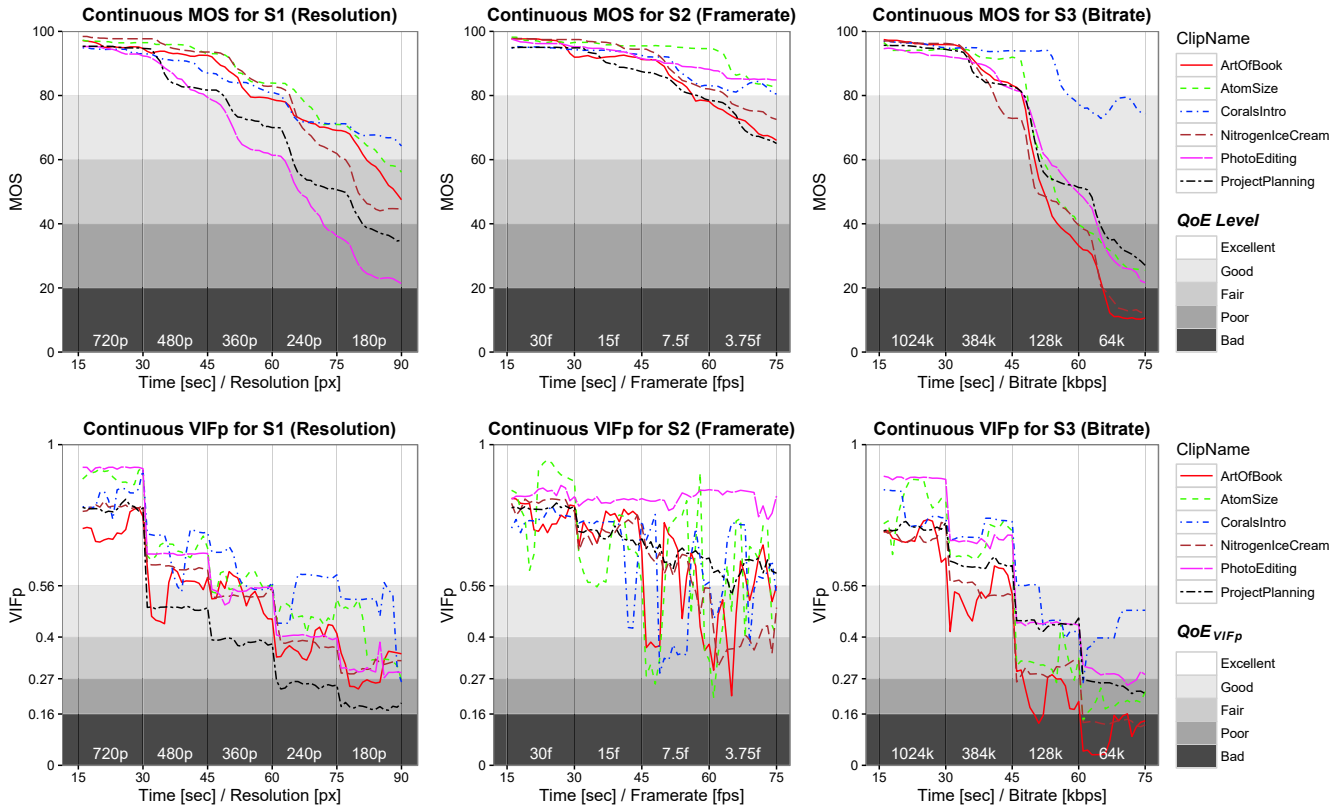

Fig. 3. Continuous MOS and VIFp values for the resolution (S1), framerate (S2) and bitrate (S3) decrease testing scenarios.

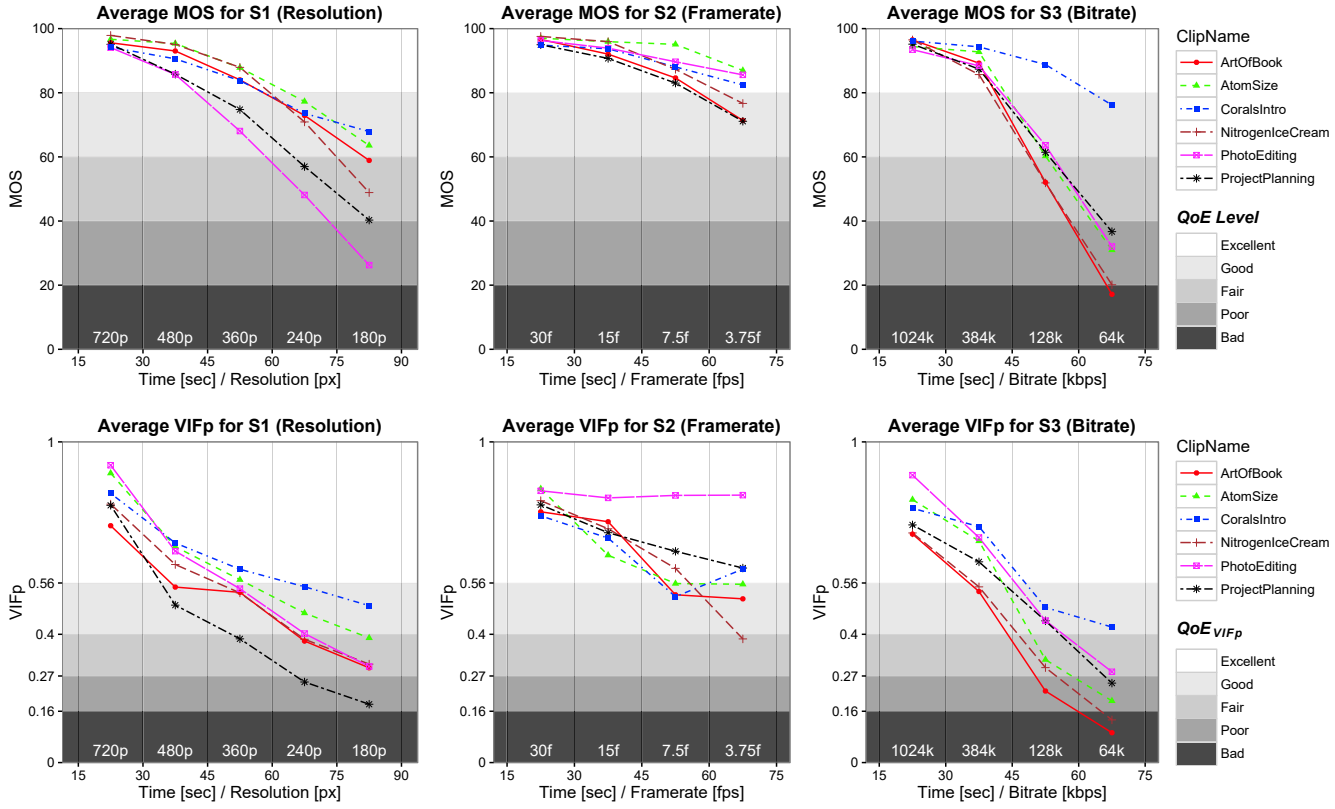

Fig. 4. Average MOS and VIFp values for each 15 second interval of the resolution (S1), framerate (S2) and bitrate (S3) decrease testing scenarios.

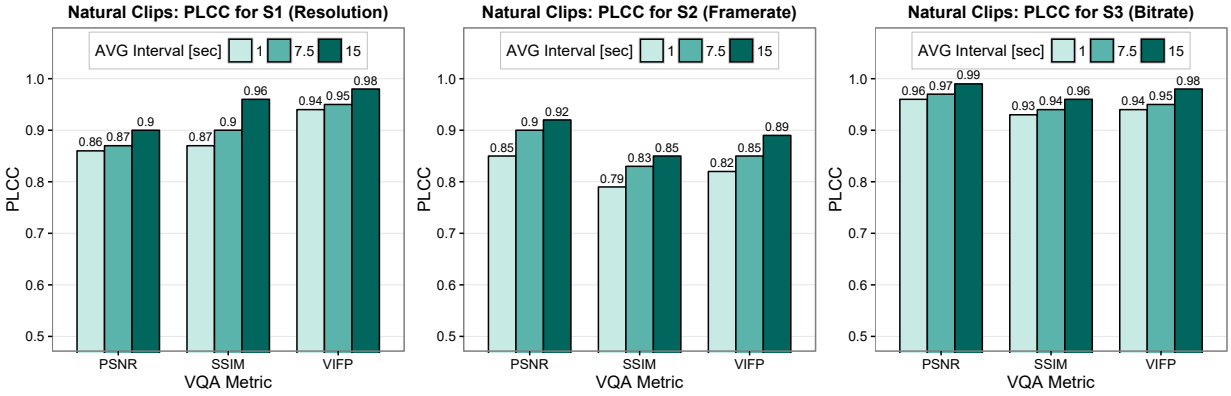

Fig. 5. PLCC quality estimation accuracy results of PSNR, SSIM and VIFp full-reference objective VQA metrics for the natural clips.

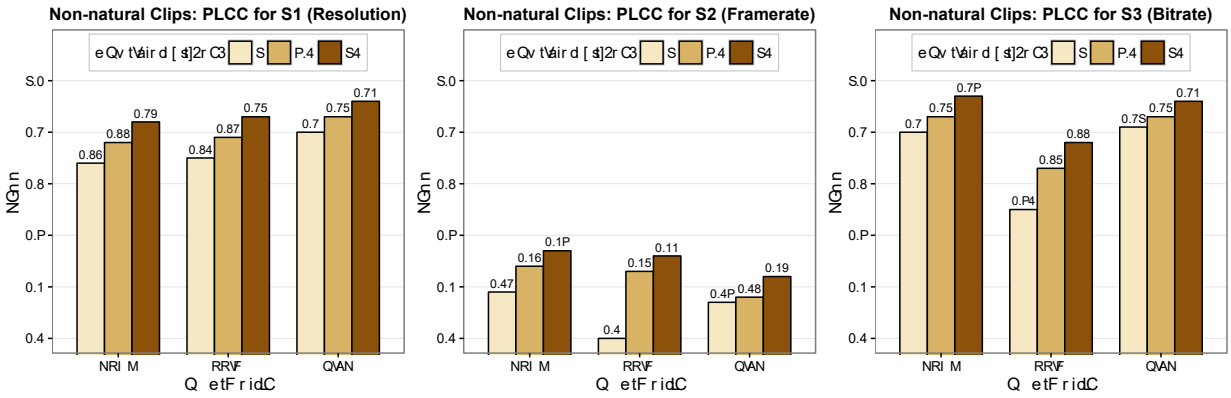

Fig. 6. PLCC quality estimation accuracy results of PSNR, SSIM and VIFp full-reference objective VQA metrics for the non-natural clips.

REFERENCES

[1] Cisco Systems, "Cisco Visual Networking Index: Global Mobile Data Traffic Forecast Update, 2015–2020 White Paper," Tech. Rep., Feb. 2016. [Online]. Available: http://www.cisco.com/c/en/us/solutions/collateral/service-provider/visual-networking-index-vni/mobile-white-paper-c11-520862.html

[2] A.-N. Moldovan, I. Ghergulescu, S. Weibelzahl, and C. H. Muntean, "User-centered EEG-based Multimedia Quality Assessment," in *8th IEEE International Symposium on Broadband Multimedia Systems and Broadcasting (BMSB 2013)*. London, UK: IEEE, Jun. 2013.

[3] S. Chikkerur, V. Sundaram, M. Reisslein, and L. J. Karam, "Objective Video Quality Assessment Methods: A Classification, Review, and Performance Comparison," *IEEE Transactions on Broadcasting*, vol. 57, no. 2, pp. 165–182, Jun. 2011.

[4] S. Winkler, "Analysis of Public Image and Video Databases for Quality Assessment," *IEEE Journal of Selected Topics in Signal Processing*, vol. 6, no. 6, pp. 616 –625, Oct. 2012.

[5] A.-N. Moldovan and C. H. Muntean, "Subjective Assessment of BitDetect - A Mechanism for Energy-Aware Multimedia Content Adaptation," *IEEE Transactions on Broadcasting*, vol. 58, no. 3, pp. 480–492, 2012.

[6] ITU-R, "BT.500-12: Methodology for the subjective assessment of the quality of television pictures," International Telecommunication Union, Geneva, Switzerland, Tech. Rep., Sep. 2009.

[7] ITU-T, "P.910: Subjective video quality assessment methods for multimedia applications," International Telecommunication Union, Geneva, Switzerland, Tech. Rep., Apr. 2008.

[8] P. Spachos, W. Li, M. Chignell, A. Leon-Garcia, L. Zucherman, and J. Jiang, "Acceptability and Quality of Experience in over the top video," in *2015 IEEE International Conference on Communication Workshop (ICCW)*, Jun. 2015, pp. 1693–1698.

[9] Z. Wang, L. Lu, and A. C. Bovik, "Video Quality Assessment Based on Structural Distortion Measurement," *Signal Processing: Image Communication, Special Issue on "Objective Video Quality Metrics"*, vol. 19, no. 2, pp. 121–132, Feb. 2004.

[10] H. Sheikh and A. Bovik, "Image Information and Visual Quality," *IEEE Transactions on Image Processing*, vol. 15, no. 2, pp. 430–444, 2006.

[11] J. Wu, W. Lin, G. Shi, and A. Liu, "Reduced-Reference Image Quality Assessment With Visual Information Fidelity," *IEEE Transactions on Multimedia*, vol. 15, no. 7, pp. 1700–1705, Nov. 2013.

[12] A. Rehman and Z. Wang, "Reduced-Reference Image Quality Assessment by Structural Similarity Estimation," *IEEE Transactions on Image Processing*, vol. 21, no. 8, pp. 3378–3389, Aug. 2012.

[13] Y. Han, Z. Yuan, and G.-M. Muntean, "No reference objective quality metric for stereoscopic 3D video," in *IEEE International Symposium on Broadband Multimedia Systems and Broadcasting (BMSB)*, 2014.

[14] C. Chen, L. Song, X. Wang, and M. Guo, "No-reference Video Quality Assessment on Mobile Devices," in *2013 IEEE International Symposium on Broadband Multimedia Systems and Broadcasting (BMSB)*, 2013.

[15] A.-N. Moldovan, I. Ghergulescu, and C. H. Muntean, "Learning Assessment for Different Categories of Educational Multimedia Clips in a Mobile Learning Environment," in *Proceedings of 25th Society for Information Technology and Teacher Education International Conference (SITE 2014)*. Jacksonville, Florida, USA: AACE, 2014, pp. 1687–1692.

[16] B. Rudiak-Gould, "AviSynth." [Online]. Available: http://avisynth.nl

[17] A.-N. Moldovan, I. Ghergulescu, and C. H. Muntean, "Performance evaluation of EMOS model for mapping-based Video Quality estimation," in *2015 9th International Symposium on Image and Signal Processing and Analysis (ISPA)*. Zagreb, Croatia: IEEE, Sep. 2015, pp. 120–125.

[18] ——, "Educational Multimedia Profiling Recommendations for Device-Aware Adaptive Mobile Learning," in *IADIS International Conference e-Learning 2014 (eL2014) (part of MCCSIS 2014)*. Lisbon, Portugal: IADIS, Jul. 2014, pp. 125–132.

[19] ——, "A Novel Methodology for Mapping Objective Video Quality Metrics to the Subjective MOS Scale," in *9th IEEE International Symposium on Broadband Multimedia Systems and Broadcasting (BMSB 2014)*. Beijing, China: IEEE, Jun. 2014, pp. 1–7.